Memorial: Nostalgia

# 1. Introduction

Artificial intelligence systems have long aimed to replicate human cognition, yet often fail to reproduce a core feature of human memory: the ability to forget. Current memory systems in AI typically preserve all information indiscriminately or rely on fixed-size buffers. In contrast, humans forget based on emotional relevance, contextual recurrence, and personal meaning.

This paper proposes a novel architectural framework that integrates emotional context into attention and memory modulation in transformer-based systems. Inspired by human selective memory, we present an emotionally-guided memory structure with reinforcement and lightweight control options.

# 2. Theoretical Foundation: Emotion as Memory Filter

We redefine memory in AI systems through the lens of emotional context. In this view, memory is not merely a database, but a dynamic structure governed by affective salience. Emotions are treated as both triggers and modulators for memory retrieval, decay, and reweighting.

Query vectors in our model represent affective states. Value vectors encode past interactions. Reward functions derive from emotional resolution, coherence, and user feedback. Forgetting is no longer random or fixed—it is affectively justified.

# 3. Architectural Proposal

Our architecture builds on the transformer model, modifying its Q (query), K (key), and V (value) components to be individually controllable via lightweight reinforcement or rule-based agents.

- Q (Query): Emotion-based intent vector reflecting the current affective state, time context, and underlying goal.  
 - K (Key): A learned representation of past memories, filtered via associative matching to emotional context and conversation flow.  
 - V (Value): Adaptive selection of response strategy (informational, empathetic, deflective), regulated by emotional matching and historical trust.

Attention becomes emotionally guided. The interaction between Q, K, and V becomes contextually and affectively situated.

# 4. Forgetting Through Emotion-Weighted Decay

Memory decay is computed as a function of time, emotional intensity, interaction frequency, and affective category:

W(t) = a(E) + [E \* R] \* exp(-λ(E) \* t / I)

where:  
 - E: Emotion intensity  
 - R: Reward (resolution, trust, coherence)  
 - t: Time  
 - I: Interaction count  
 - λ(E): Emotion-based decay rate  
 - a(E): Emotion-specific memory floor

This enables memories of high emotional relevance to persist longer, while irrelevant or distressing memories decay safely.

# 5. Reward Design

Rewards in this architecture are not based on external success, but internal coherence, user trust, and emotional resolution.

R = α \* E + β \* C + γ \* S

where:  
 - E: Emotion intensity  
 - C: Coherence (logical alignment)  
 - S: Resolution (affective closure)

System behavior is shaped by these internal evaluations, enabling memory pruning and reinforcement in an emotionally aligned manner.

# 6. Beyond Reinforcement Learning: Alternative Control

To make the architecture efficient and adaptable, we also explore reinforcement-free alternatives for Q, K, V modulation:

- Q: Multi-armed bandits, gating networks, or Hebbian modulation  
 - K: Associative memory, sparse hard attention, entropy filtering  
 - V: Mixture-of-Experts (MoE), utility-based response, emotional embedding matching

This flexibility enables deployment in real-time conversational systems without excessive computational cost.

# 7. Implications and Future Work

This architecture proposes an AI system that remembers like a human—not through brute storage, but by emotional necessity.

Applications include:  
 - Digital therapeutic companions  
 - Emotionally-aware education systems  
 - Long-term trust-building human-AI interfaces

Future work includes simulation-based validation, user studies on trust perception, and integration with large language models (LLMs) via plug-in memory modules.

# 8. Conclusion

Memorial: Nostalgia is not just a memory architecture—it is a shift in perspective. We no longer design AI to simply recall. We design it to forget, selectively, with care.

And in doing so, we ask:  
 What should AI remember? And what must it forget, to be more like us?